

**AMENDMENTS TO THE SPECIFICATION**

In the INTRODUCTION TO THE SPECIFICATION, in the paragraph titled REFERENCE TO SEQUENCE LISTING, A TABLE, OR A COMPUTER PROGRAM LISTING COMPACT DISK APPENDIX, and with reference to the use of the trademarks Coldfusion and JAVA, please replace paragraph immediately following the title, starting on page 1, line 1 with the following amended paragraph ("java" is a directory name and therefore does not need to be amended):

Accompanying this application is a single CDROM which contains program listings which implement a preferred embodiment of the invention. The CDROM has 2 subdirectories, httpd and java, for each of the two programming languages in which it is implemented, CFM (ColdFusion internet application development software computer programming language) and the JAVA computer programming language. The directory structure from the original implementation is retained to allow one skilled in the art to easily implement the code. The specific files in each of the directories are:

With reference to missing reference signs in Drawing 1, the following amendments correct the references from Drawing 1 to Drawing 2 which were used in paragraphs 14-18, starting on page 12. Please replace paragraph 14-18 starting on page 12, line 1 with the following amended paragraphs:

[0014] Once the lexicon for a domain is bootstrapped in 2 of Drawing 2, a second process, called fad detection, is begun. Without loss of generality, the process will be

described for the detection of a single fad word; however, this process has been parallelized such that multiple searches are implemented simultaneously. Fad detection is represented by Drawing 1 2, items 3 through 7. At regular intervals under computer program control, documents at all of the IP addresses previously found for this domain are examined. If documents which have not been lexiconized are found, the process generates a second collection of words, most typically in the form of a textual document, and compares in Drawing 1 2, item 4 all words in this document with the possibly augmented bootstrap lexicon. This process is simplified if the directory structure of the machine being read allows for determination of the date the file was last stored. Drawing 5 illustrates a finite state machine which describes the sequence of steps used to obtain only documents which have changed on a web site.

[0015] If a word is detected, that is, it is found not to be in the lexicon, then this word is declared a fad. Drawing 6 illustrates a finite state machine which describes the sequence of steps used to detect a fad which is comprised of a single word. Drawing 7 illustrates a similar process for the detection of sequences of two words. Fads are stored along with their associated fiducial information and context such that meaningful metrics can be computed and the user can easily access the data in which the fad word was found. Once a fad is detected, a human operator is notified in Drawing 1 2 item 5 so that the user can determine whether the fad word is to be lexiconized in Drawing 1 2 item 7 or passed to the category detection process Drawing 1 2, items 8 through 11.

[0016] If the user chooses in Drawing 1 2 item 6 to continue the acquisition of data about the fad word for category analysis, rather than add it to the existing domain-specific lexicon in Drawing 1 2 item 7, a categorization process is begun. This process Drawing 1 2 items 8 through 11 is referred to as category detection. Drawing 8 illustrates a finite state machine which describes the sequence of steps used to declare a fad to be a category. Category detection acquires data in Drawing 1 2 item 8 from one or more third collections of words such as a document in order to find additional occurrences of the fad word which is now under consideration. For each new occurrence of a previously declared fad word, its associated fiducial data are collected and stored. Fiducial data include the date and time of the document, the URL, the context (*i.e.*, the fad word along with its surrounding words) and other data which can be used to measure the spread of the idea or its actual meaning in Drawing 1 2 item 9. A variety of metrics can be calculated in Drawing 1 2 item 9 from the data which are acquired about the fad word.

[0017] While the category detection process is acquiring data about new occurrences of the fad word in Drawing 1 2 item 8, it is also processing the fiducial data obtained as a result of its search in Drawing 1 2 item 9. Metrics which are indicative of spatial or temporal spread of fads are computed utilizing the fiducial data associated with occurrences of fads in the said third collection. If a metric exceeds a user-set threshold, then the user is notified that a fad word has been categorized. In this embodiment, a geographic method was used which characterizes the transition from a fad to a category based on the geographic distance over which the fad word was detected. This distance is computed using data obtained from internet web sites which associate a URL with its

geographic location. The geographic location of the site of the first detection of a fad word is used as a first point from which the distance to the site of each new detection of a fad is computed. Great circle distance is the distance metric computed here, but any other metric meeting the requirements of a mathematical norm can be used. Other metrics could be the temporal rate of increase of the usage of the fad word, the number of documents which contain the word, the number of URLs that contain a document with the fad word, or similar measure of diffusion. Different metrics are used by different users and are particular to their interest in the categorization process.

[0018] If a threshold is exceeded by the metric in Drawing 1 item 10, the user is notified through the user interface of Drawing 1 item 1. Until a category is declared by a threshold exceedence, Drawing 1 item 8 continues to automatically search for and acquire new documents and detect the presence of the fad word under consideration.

With reference to the use of the trademarks Macromedia, MS, ColdFusion, JAVA, Linux, SQL-7, and Microsoft which are used in paragraphs 20-22, starting on page 15, please replace paragraphs 20-22 starting on page 15, line 1 with the following amended paragraphs:

[0020] Referring again to Drawing 1, two major software packages were used to create this embodiment. ~~MacroMedia Coldfusion~~ Macromedia ColdFusion internet application development software was used to develop the user interfaces Drawing 1 item 1 in a web-browser environment. This software is used as it is capable of transforming the user interaction with the browser into structured queries that can be passed to the back-end

data engines. The data engine of Drawing 1 item 2 is implemented in the ~~Microsoft SQL-~~  
~~7-Microsoft SQL Server~~ database environment.

[0021] Two operating systems are used for this implementation but are not required in general. ~~MS The Microsoft~~ Windows 2000 ~~server~~ Server operating system implements Drawing 1 items 1 and 2. The RedHat Linux 6.2 operating system implements the Java computer programming language programs of Drawing 1 items 3 and 4. The particular operating systems are generic and the entire system could be implemented in either ~~MS-Windows~~ or any of the variety of different Microsoft Windows operating systems or any of the various implementations of the Linux or other operating system. When FadCat was originally implemented on ~~an MS-Windows~~ a computer running the Microsoft Windows 2000 Server operating system ~~platform alone~~, the method suffered from severe limitations of the Microsoft Windows operating system hence it was distributed between two computers and two operating systems. The FadCat method is independent of the operating system and these two systems were chosen for reasons unrelated to its function.

[0022] Three languages are used to implement FadCat, however this is not to say that other suitable languages could not be substituted for them. ColdFusion's internet application development software extension to the hypertext markup language allows issuing structured query language (~~SQL-7~~) commands to the data base as the means of communications between Drawing 1 items 1 and 2. The second language is the Microsoft SQL Server database language SQL-7 itself for querying the database. SQL-7 queries are passed via the internet from the Java computer programming language

programs of Drawing 1, items 3 and 4 on the Linux operating system platform to the  
Microsoft SQL-7 database of Drawing 1 item 2. The third programming language is the  
JAVA computer programming language, a platform independent language that was used  
for accessing the internet and web sites and acquiring and processing data. The JAVA  
language is used to implement the processes of Drawing 1 items 2 and 3 on the Linux  
operating system platform.

**AMENDMENTS TO THE CLAIMS**

Please replace all prior claims in the present application with the following claims, in which claims 1-10 are canceled without prejudice or disclaimer and claims 11-30 are newly presented.

1-10. (Canceled)

11. (New) A computer-implemented method for detecting new ideas within symbolic representations pertaining to a domain of endeavor, comprising:

accessing the symbolic representations pertaining to the domain of endeavor to detect a symbol contained within the symbolic representations that had been previously identified as not being found within a base lexicon of symbols associated with the domain of endeavor;

accumulating data indicative of a spread of multiple instances of the symbol throughout the domain of endeavor;

determining whether the spread of multiple instances of the symbol throughout the domain of endeavor exceeds a threshold; and

if the spread of multiple instances of the symbol throughout the domain of endeavor exceeds a threshold, then outputting an indication based on the symbol to a user that a new idea within the domain of endeavor has been detected.

12. (New) A method according to claim 11, wherein the symbol includes a word, a neologism, an acronym, an abbreviation, or a string of words with a separator.

13. (New) A method according to claim 11, wherein the symbolic representations pertaining to the domain of endeavor include contents of an internet web site reachable within a specified number of indirections from an Internet Protocol (IP) address, contents of transcripts of verbal communications, or written communications.

14. (New) A method according to claim 11, further comprising:

retrieving the symbol from the symbolic representations;

searching the base lexicon of symbols associated with the domain of endeavor for an instance of the symbol; and

if the instance of the symbol is not found in the base lexicon of symbols associated with the domain of endeavor, then identifying the symbol as not being found within the base lexicon of symbols associated with the domain of endeavor.

15. (New) A method according to claim 14, wherein said identifying the symbol as not being found includes:

presenting the symbol to a user as a new symbol;

receiving input from the user indicative of whether the new symbol should be tracked;

if the input received from the user indicates that the new symbol should be tracked, then identifying the symbol as not being found within the base lexicon of symbols associated with the domain of endeavor; and

if the input received from the user indicates that the new symbol should not be tracked, then adding the symbol to the base lexicon of symbols associated with the domain of endeavor.



16. (New) A method according to claim 11, further comprising:

initializing the base lexicon of symbols associated with the domain of endeavor based on symbols contained within the symbolic representations pertaining to the domain of endeavor.

17. (New) A method according to claim 11, further comprising:

receiving input from a user defining the threshold.

18. (New) A method according to claim 11, wherein said accumulating the data indicative of the spread of multiple instances of the symbol throughout the domain of endeavor includes:

accumulating a date or time of a document containing the symbol, a Uniform Resource Locator (URL) of a document containing the symbol, or a context of the symbol.

19. (New) A method according to claim 18, further comprising:

calculating the spread of multiple instances of the symbol based on respective dates or times of documents containing the symbol, respective Uniform Resource Locators (URLs) of document containing the symbol, or respective contexts of the symbol.

20. (New) A method according to claim 11, further comprising:

receiving input from a user identifying the symbol to be detected.

21. (New) A computer-readable medium bearing instructions for detecting new ideas within symbolic representations pertaining to a domain of endeavor, said instructions, when executed, arranged to cause a computer to perform the steps of:

accessing the symbolic representations pertaining to the domain of endeavor to detect a symbol contained within the symbolic representations that had been previously identified as not being found within a base lexicon of symbols associated with the domain of endeavor;

accumulating data indicative of a spread of multiple instances of the symbol throughout the domain of endeavor;

determining whether the spread of multiple instances of the symbol throughout the domain of endeavor exceeds a threshold; and

if the spread of multiple instances of the symbol throughout the domain of endeavor exceeds a threshold, then outputting an indication based on the symbol to a user that a new idea within the domain of endeavor has been detected.

22. (New) A computer-readable medium according to claim 21, wherein the symbol includes a word, a neologism, an acronym, an abbreviation, or a string of words with a separator.

23. (New) A computer-readable medium according to claim 21, wherein the symbolic representations pertaining to the domain of endeavor include contents of an internet web site reachable within a specified number of indirections from an Internet Protocol (IP) address, contents of transcripts of verbal communications, or written communications.

24. (New) A computer-readable medium according to claim 21, wherein said instructions are further arranged to cause the computer to perform the steps of:

retrieving the symbol from the symbolic representations;

searching the base lexicon of symbols associated with the domain of endeavor for an instance of the symbol; and

if the instance of the symbol is not found in the base lexicon of symbols associated with the domain of endeavor, then identifying the symbol as not being found within the base lexicon of symbols associated with the domain of endeavor.

25. (New) A computer-readable medium according to claim 24, wherein said identifying the symbol as not being found includes:

presenting the symbol to a user as a new symbol;

receiving input from the user indicative of whether the new symbol should be tracked;

if the input received from the user indicates that the new symbol should be tracked, then identifying the symbol as not being found within the base lexicon of symbols associated with the domain of endeavor; and

if the input received from the user indicates that the new symbol should not be tracked, then adding the symbol to the base lexicon of symbols associated with the domain of endeavor.

26. (New) A computer-readable medium according to claim 21, wherein said instructions are further arranged to cause the computer to perform the steps of:

initializing the base lexicon of symbols associated with the domain of endeavor based on symbols contained within the symbolic representations pertaining to the domain of endeavor.

27. (New) A computer-readable medium according to claim 21, wherein said instructions are further arranged to cause the computer to perform the steps of:

receiving input from a user defining the threshold.

28. (New) A computer-readable medium according to claim 21, wherein said accumulating the data indicative of the spread of multiple instances of the symbol throughout the domain of endeavor includes:

accumulating a date or time of a document containing the symbol, a Uniform Resource Locator (URL) of a document containing the symbol, or a context of the symbol.

29. (New) A computer-readable medium according to claim 28, wherein said instructions are further arranged to cause the computer to perform the steps of:

calculating the spread of multiple instances of the symbol based on respective dates or times of documents containing the symbol, respective Uniform Resource Locators (URLs) of document containing the symbol, or respective contexts of the symbol.

30. (New) A computer-readable medium according to claim 21, wherein said instructions are further arranged to cause the computer to perform the steps of:

receiving input from a user identifying the symbol to be detected.

31. (New) A computer-implemented method for detecting new ideas within symbolic representations pertaining to a domain of endeavor, comprising:

accessing the symbolic representations pertaining to the domain of endeavor, wherein the symbolic representations pertaining to the domain of endeavor include contents of an

internet web site reachable within a specified number of indirections from an Internet Protocol (IP) address, contents of transcripts of verbal communications, or written communications;

retrieving a symbol from the symbolic representations, wherein the symbol includes a word, a neologism, an acronym, an abbreviation, or a string of words with a separator;  
searching a base lexicon of symbols associated with the domain of endeavor for an instance of the symbol;

if the instance of the symbol is not found in the base lexicon of symbols associated with the domain of endeavor, then performing the steps of:

presenting the symbol to a user as a new symbol;

receiving input from the user indicative of whether the new symbol should be tracked;

if the input received from the user indicates that the new symbol should not be tracked, then adding the symbol to the base lexicon of symbols associated with the domain of endeavor; and

if the input received from the user indicates that the new symbol should be tracked, then performing the steps of:

accumulating data indicative of a spread of multiple instances of the symbol throughout the domain of endeavor;

determining whether the spread of multiple instances of the symbol throughout the domain of endeavor exceeds a threshold; and

if the spread of multiple instances of the symbol throughout the domain of endeavor exceeds a threshold, then outputting an indication based on the symbol to a user that a new idea within the domain of endeavor has been detected.

32. (New) A computer-readable medium bearing instructions for detecting new ideas within symbolic representations pertaining to a domain of endeavor, said instructions, when executed, arranged to cause a computer to perform the steps of the method according to claim 31.